第 40 卷　第 2 期
2022 年 6 月

天　文　学　进　展
PROGRESS IN ASTRONOMY

Vol. 40,　No. 2
June., 2022

doi: 10.3969/j.issn.1000-8349.2022.02.07

应用于引力波探测的深度学习网络结构复杂度研究

马存良[1]，詹　超[1]，嘉明珍[1]，贺观圣[2]，李伟军[2]，易见兵[1]

(1. 江西理工大学，赣州 341000；　2. 南华大学，衡阳 421200)

摘要: 深度学习用于引力波探测是近几年的研究热点。匹配滤波法可以看作模板存储于卷积核参数中的单卷积层的神经网络，通过加深模型的深度可以在参数大大减少的同时得到相似的探测效果。对不同的卷积核大小、卷积核的数量 (模型的宽度)、卷积层的数量 (模型的深度) 的深度学习引力波探测模型进行了讨论。另外，对全连接层前采用批量归一优化 (batch normalization, BN) 模型的探测效果进行了研究，发现单卷积层的模型在加入 BN 后的探测精度由 50% 左右提高到了 90% 以上。研究结果为匹配滤波模板数量的压缩提供了潜在的新方法，匹配滤波后通过BN 层和全连接层也许能够大大减少匹配模板数量。

关　键　词: 双黑洞并合；引力波探测；深度学习

中图分类号: P142.8+4　　　**文献标识码:** A

1　引力波探测的研究背景

2015 年 9 月 14 日，激光干涉引力波天文台 (Laser Interferometer Gravitational-wave Observatory, LIGO) 在人类历史上首次直接探测到引力波事件，代号 GW150914，这个引力波事件来自 13 亿光年外的两个黑洞并合。引力波的成功探测直接验证了广义相对论中引力波存在的预言，是人类科学史上的伟大胜利，打开了全新的宇宙观测窗口——引力波天文学[1]。2017 年 8 月 17 日，人类历史上第一次直接观测到双中子星并合释放的引力波信号GW170817[2]，在检测到并合信号约 1.7 s 后，费米伽马射线暴监视系统检测到伽马射线暴[3–5]，在接下来的数天内分布在世界各地的多个天文观测站独立检测到多个频段的电磁信号[6]，这是第一次从单一天体物理源对引力波和电磁辐射进行明确一致的观测，标志着多信

收稿日期: 2021-08-02；　**修回日期:** 2021-10-08

资助项目: 国家自然科学基金 (11847143)；江西省自然科学基金 (20181BAB202004)；江西省研究生创新基金 (YC2020-S469)

通讯作者: 马存良，mcl@jxust.edu.cn

使天文学时代的到来。

地基激光干涉仪探测器的灵敏探测频段是 $10 \sim 10^3$ Hz，主要探测双中子星和恒星级双黑洞并合释放的引力波信号。除了地面引力波探测器外，空间引力波探测器也在计划中。空间激光干涉引力波探测器探测引力波的频段为 $10^{-4} \sim 1$ Hz，主要探测银河系内大量的双白矮星系、极端和中等质量比双星系统等。空间引力波探测器 (Laser Interferometer Space Antenna, LISA) 被列为欧洲航空局的 L3 计划[7]，中国天琴[8]和太极[9]也在规划当中。2019 年 8 月 31 日，微重力技术实验卫星“太极 1 号”在酒泉卫星发射中心成功发射，各项测试结果正常。2019 年 12 月 20 日，天琴计划的第一颗技术实验卫星，“天琴 1 号”在太原卫星发射中心成功发射升空。目前引力波探测实验卫星 LISA Pathfinder、太极 1 号、天琴 1 号都运行非常成功，预计 21 世纪 30 年代空间引力波探测器将会运行，多频段引力波探测时代将会到来。

引力波探测需要数学、物理学、计算机科学等多学科协同攻关，并且交叉融合[10]。LIGO 的精度可以达到 10^{-22} 量级，在测量的信号中，背景噪声通常大于引力波信号，要通过数据分析流水线来搜索信号中的瞬态引力波模式。引力波数据分析流水线包括 PyCBC[11], GstLAL[12], SPIIR[13]等，在加州理工学院、麻省理工学院、汉福德观测站、利文斯顿观测站等约 7 万 CPU 核心和数千块高性能计算显卡上运行。

中国、澳洲和美国的科研工作者共同开发的匹配滤波算法 SPIIR[13]，创造性地由并行时间平移的一阶 IIR 滤波器构造匹配滤波器，把模板库中的模板映射到大量的时间平移 IIR 滤波器参数，使算法更易于并行和扩展。引力波数据分析流水线有离线和在线两种工作模式，后者的数据处理延时可以达到分钟量级。多信使天文学的顺利开展需要更低的计算延时，实时探测推断的天空方位可以与电磁探测望远镜共享，以便快速寻找相关方位的电磁信号。

在 LIGO 数据处理流水线中，已经融入机器学习方法。LIGO 通过约 5 000 个快数据通道和约 20 万个慢数据通道的分析得到信号质量的好坏，频率为 60 Hz 的交流电源的非线性耦合瞬态噪声已经可以通过机器学习法线下去除[14]。我们相信机器学习方法也必将应用于引力波的探测、定位和波源参数反演等。深度学习有望集成到引力波数据分析流水线中，替代匹配滤波的低延迟分析，助力引力波探测实时分析。

深度学习在计算机视觉、声音信号处理、医疗诊断等方面的应用已经取得显著的成效，在引力波数据处理方面的应用还在探索中。2018 年，Gabbard 等人[15]、George 和 Huerta[16]几乎同时发现，深度学习在引力波探测方面可以达到匹配滤波的精度。与匹配滤波相比，深度学习的计算效率大大提高，有望应用于瞬态引力波模式的实时检测。后来，George 和 Huevta[17]将深度学习模型应用于真实的 LIGO 噪声信号，发现其与仿真噪声信号一致。近几年许多工作对深度学习引力波探测进行了研究，这些研究可以划分为模型优化、应用拓展、应用细化和机理研究。

2019 年，Gebhard 等人[18]使用空洞卷积神经网络对引力波探测进行了研究，他们发现空洞卷积神经网络可以有效检测致密双星并合引力波信号的有无。2020 年，Wang 等人[19]在卷积神经网络基础上加入匹配感知层，得到一种新的网络结构，该网络的感知层利用几十个

波形模板模拟匹配滤波的过程。Li 等人[20]通过小波分解对信号进行预处理，基于 Gabbard 模型对参数进行调整，得到优化的模型。Luo 等人[21]研究发现，通过对模型中 dropout 层的改进可以增加探测精度。Deighan 等人[22]采用遗传算法对 George 和 Huerta 的模型进行优化，得到一个参数数量为原模型的 78% 但精度提高 11% 的模型。2021 年，Xia 等人[23]对卷积神经网络结构在引力波探测方面的泛化能力做了研究，他们以 George 和 Huerta 的模型为基础，通过 dropout 和 BN 优化了模型的性能，并对输入输出的统计特性进行研究后得出，模型对不同的输入参数空间具有不同的输出统计特性。Huerta 团队集成两个卷积神经网络得到一个新模型，该模型可以对引力波探测网络的数据进行并发分析[24]。

除了应用于致密双星引力波信号探测外，科研人员应用深度学习对波源参数的预测与定位[25–28]、引力波去噪[29, 30]、连续引力波探测[31]、星体坍缩引力波探测[32, 33]等进行了研究。2019 年，Huerta 团队构建深度模型预测双星的质量、自旋等参数，得到的结果与贝叶斯分析方法一致，但是分析时间大大降低，模型预测时间在单块 Tesla V100 GPU 上为 2 ms[25]。2020 年，该团队又将应用于语音信号处理的 WaveNet 拓展到引力波去噪[30]。Fan 等人[26]创新性地将引力波探测和波源参数估算融合在一个模型中，并且研究了探测精度随观测站数量的变化。2021 年重庆大学李瑾团队采用机器学习对脉冲星计时阵列信号进行分析，研究了纳赫兹波段引力波的探测问题[34]。

本文以 LIGO O1 真实噪声数据为基础，研究了不同网络结构复杂度的深度模型对引力波探测的影响；第 2 章回顾了匹配滤波和卷积神经网络的数学表达式，发现两者的一致性；第 3 章和第 4 章分别介绍了应用于模型训练和测试的数据集和模型的结构；第 5 章对实验结果进行了展示和分析。

2　匹配滤波与卷积神经网络的一致性

2.1　匹配滤波法

在平稳高斯可加背景噪声假设下，可以得到最优线性引力波探测方法——匹配滤波法。假设 $s(t)$ 为 LIGO 获取的应变信号，$h_i(t) \in T$ 为匹配模板，其中 T 为模板库，则模板 $h_i(t)$ 对应的匹配滤波输出为：

$$z_i(t) = \int_{-\infty}^{\infty} \frac{\tilde{s}(f)\tilde{h}_i^*(f)}{S_{dn}(f)} \mathrm{e}^{j2\pi ft}\mathrm{d}f \quad , \tag{1}$$

其中，$\tilde{s}(f)$ 和 $\tilde{h}_i^*(f)$ 分别为 $s(t)$ 和 $h_i(t)$ 的傅里叶变换。$S_{dn}(f)$ 为信号的双边功率谱，与单边功率谱 $S_n(t)$ 的关系为：

$$S_n(f) = \begin{cases} 0 & , \ f < 0 \\ 2S_{dn}(f), & f > 0 \end{cases} \quad , \tag{2}$$

设

$$x(t) = \int_{-\infty}^{\infty} \frac{\tilde{s}(f)}{S_n(f)} \mathrm{e}^{j2\pi ft}\mathrm{d}f \quad , \tag{3}$$

则模板 $h_i(t)$ 对应的匹配滤波输出为：

$$z_i(t) = \int x(\tau)h_i(t-\tau)\mathrm{d}\tau \quad , \tag{4}$$

LIGO 获取的应变数据为离散时间信号 $s[n] = s(nT)$，其中 T 为采样周期，对上式进行离散化后得到：

$$z_i[n] = T\sum_m x[m]h_i[m-n] \quad , \tag{5}$$

其中，$h_i[n] = h_i(nT)$。则匹配滤波的输出为：

$$z[n] = \max_i (z_i[n]) \quad . \tag{6}$$

2.2 卷积神经网络中的卷积层

卷积神经网络由输入层、输出层和若干隐藏层构成。隐藏层可以由卷积、池化、全连接层等层层堆叠实现，每一层都可以看成具有输入输出的系统。卷积层可以由卷积核的大小、数量等确定，假设卷积层的输入为 $x[n]$，第 i 个卷积核为 $w_i[n]$，则其产生输出[35]：

$$y_i[n] = \sum_m x[m]w_i[n-m] \quad , \tag{7}$$

令 $r_i(t) = h_i(-t)T$，式 (5) 变为：

$$z_i[n] = \sum_m x[m]r_i[n-m] \quad . \tag{8}$$

对比式 (7) 和式 (8)，我们可以得出，匹配滤波可以看作具有一层卷积层 (卷积核参数构成了匹配模板库) 的卷积神经网络模型。

2.3 模型宽度和深度研究的必要性

我们把每层卷积核的数量称为模型的宽度，卷积层的层数称为模型的深度。匹配滤波属于宽度很宽，深度只有一层的卷积神经网络。信号与模板直接匹配感知的匹配滤波法计算效率低。已有的深度学习引力波探测模型，通过提高模型的深度来降低模型参数数量，进而提高计算效率。按照当前 LIGO 设计灵敏度，引力波探测模型的虚警概率必须非常低，低至万分之一才能应用于引力波探测。在应用于在线实时探测前，必须对深度学习引力波探测方法的优化进行更加全面的研究。卷积神经网络宽度和深度的平衡能否兼顾效率和可靠性，仍然需要研究。本文对深度模型的宽度和深度在引力波探测中的性能进行了研究。

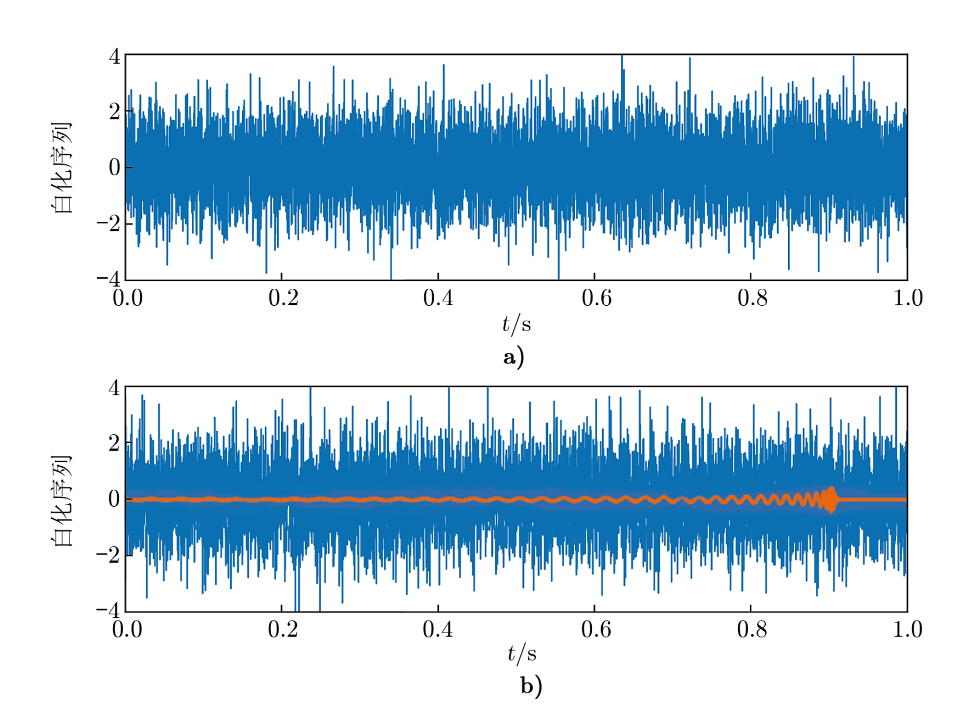
3 数据集

文中的数据集通过开源程序 ggwd 生成[18]。数据集包括两类，一类仅包含 LIGO 真实噪声，另一类包含噪声加引力波。噪声从 H1 数据中随机获取 (不包括探测事件)，引力波

信号通过 SEOBNRv4 波形近似方法生成。每条训练数据的时间长度为 1 s，采样频率为 4 096 Hz，数据集中引力波波源参数和信噪比参数范围示于表 1。图 1 展示了两类数据示例，Class 0 仅包含探测器噪声，Class 1 为探测器噪声注入引力波信号。在使用 SEOBNRv4 模型进行双黑洞并合波形仿真时，我们把引力波源距离定为恒定值 100 Mpc。但是这个波源距离在最终的引力波加噪声合成信号中失去意义，因为双黑洞并合引力波信号在注入到背景噪声前，根据预定义好的信噪比进行了缩放处理。

表 1 数据生成参数

参数名称	数据分布
双黑洞并合质量	$(5\sim80)M_\odot$ 内均匀分布
并合黑洞自旋分量	$0\sim0.998$ 之间均匀分布
赤经和赤纬	对空间方位均匀采样
极化角	$0\sim2\pi$ 之间均匀分布
相位和倾角	从球体上的均匀分布中联合采样
注入信噪比	$5\sim20$ 之间均匀分布



注：a) Class 0，只包含噪声的信号；b) Class 1，蓝色为噪声与致密双星并合引力波叠加的信号，橙色为引力波信号。图中两类信号中的噪声不同。

图 1 白化的应变信号

4 深度学习引力波探测模型

本文使用的深度学习引力波探测模型包括输入模块、隐藏模块、输出模块。输入模块接收白化的应变数据，输出通过全连接和 Softmax 激活函数获得二分类的值。隐藏模块通过基本单元的堆积构成，基本单元通过卷积、ReLU 激活、最大池化构建，如图 2 所示。本文中讨论的模型的网络复杂度由基本单元中卷积核的大小、卷积核的数量和模型中基本单元的数量共 3 个参数确定。隐藏层中基本单元的数量可以确定模型的深度，基本单元中卷积核的数量可以确定模型的宽度。另外，我们还对最后一层基本单元通过 BN 优化进行了研究。图 3 展示了具有三个基本单元构成的引力波探测模型的系统框图。表 2 列出了本文研究模型的参数取值。

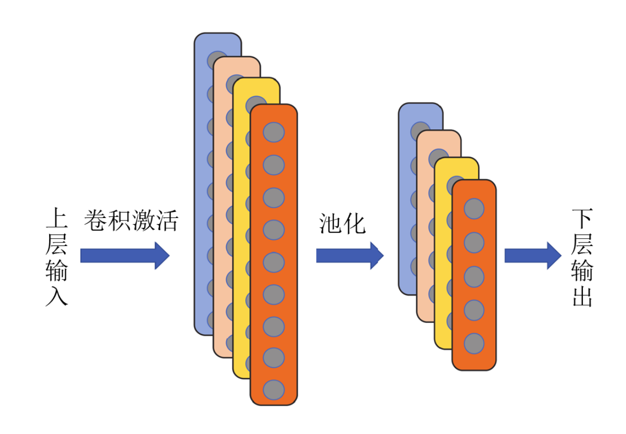
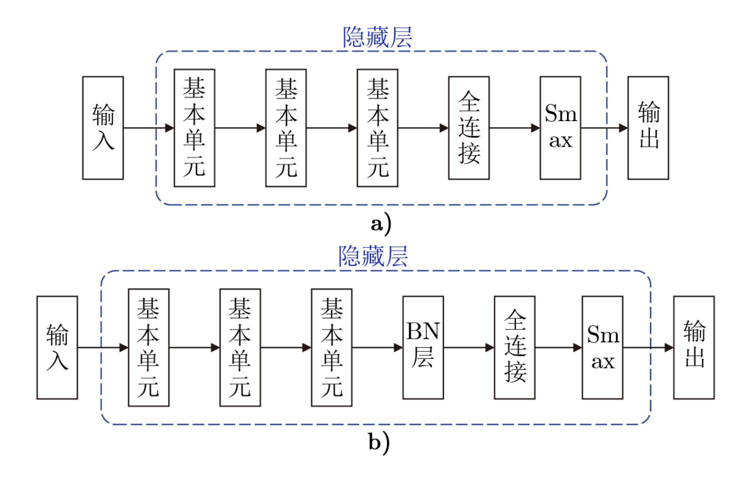


图 2 隐藏层基本模块示意图



注：a) 无 BN 优化；b) 有 BN 优化。

图 3 三个基本模块构成的引力波探测模型的系统框图

表 2 模型参数取值

参数	模型深度	模型宽度	卷积核大小
取值	1, 2, 3, 4, 5, 6, 7, 8	8, 16, 32	8, 16

5 实验结果

深度学习引力波探测模型不但与模型结构有关，而且与训练集的规模有关。本文中所有实验通过四种规模的训练集训练，分别是 4800, 20 000, 35 000, 50 000，模型的最优深度随训练集规模的变化如表 3 和表 4 所示。当卷积核大小为 8 和 16 时，最优模型的训练规模都是 50 000，所以本文后续内容基于规模为 50 000 的训练集训练得到的模型展开研究。

表 3 卷积核大小为 8 时模型的最优深度随训练集规模和模型宽度的变化

训练集规模	模型宽度					
	8		16		32	
	最优深度	准确率	最优深度	准确率	最优深度	准确率
4 800	6	91.58%	5	92.49%	3	92.83%
20 000	7	94.39%	6	94.09%	5	94.02%
35 000	7	93.43%	5	93.56%	4	93.26%
50 000	6	93.43%	4	94.86%	3	94.87%

表 4 卷积核大小为 16 时模型的最优深度随训练集规模和模型宽度的变化

训练集规模	模型宽度					
	8		16		32	
	最优深度	准确率	最优深度	准确率	最优深度	准确率
4 800	6	91.49%	5	92.99%	4	92.16%
20 000	6	94.23%	3	93.83%	4	94.02%
35 000	7	93.43%	5	93.56%	4	93.26%
50 000	6	94.17%	5	94.58%	4	95.10%

5.1 模型深度和宽度对引力波探测的影响

调整卷积核大小、模型宽度、模型深度构建深度网络模型，对模型进行训练、测试，表 5 和表 6 列出测试结果。通过实验我们可以看出，模型每层卷积核数量越多 (宽度越宽)，最优深度就越小。例如，当卷积核大小为 8 时，模型宽度为 8 的最优深度为 6，模型宽度为 32 的最优模型深度仅为 3。这说明可以通过综合对模型的深度和宽度优化来得到更优的探测模型。

5.2 BN 优化对引力波探测的影响

另外，我们还对最后一层基本单元后加入 batch normalization (BN) 进行了研究。加入 BN 层后，研究了不同深度模型的引力波探测能力。图 4 展示了加入 BN 前后引力波探测准

表 5 卷积核大小为 8 时不同深度和宽度模型引力波探测精度 (H1 训练集规模为 50 000)

	宽度		
深度	8	16	32
1	49.33%	50.66%	50.66%
2	50.66%	49.33%	49.33%
3	92.94%	93.88%	94.87%
4	92.85%	94.86%	94.32%
5	93.01%	94.63%	94.85%
6	93.43%	94.73%	93.14%
7	93.07%	94.10%	94.60%
8	93.18%	94.08%	94.22%

表 6 卷积核大小为 16 时不同深度和宽度模型引力波探测精度 (H1 训练集规模为 50 000)

	宽度		
深度	8	16	32
1	50.66%	50.66%	49.33%
2	49.33%	49.33%	50.66%
3	92.53%	92.24%	94.52%
4	91.93%	93.65%	95.10%
5	93.61%	94.58%	94.85%
6	94.17%	94.27%	94.73%
7	86.36%	94.02%	94.42%
8	93.80%	94.07%	94.35%

确率随模型深度的变化。我们发现添加 BN 层对深度为 1 和 2 的引力波探测模型的探测精度有很大的提升，如当模型深度为 1 时，探测精度由 51% 提升到 94%。由本文第 2 章分析可知，深度为 1 的卷积神经网络可以看作匹配滤波法后接全连接层，本实验的效果提供了优化匹配滤波法的新思路，也许能够通过 BN 层和全连接层改进匹配滤波，保持探测性能的同时使模板数量大大减小，达到实时探测。图 5 展示了加入 BN 前后不同深度的模型在测试集上的 ROC 曲线。从图 4 和图 5 的综合分析，我们发现 BN 在深层特征优化方面效果欠佳，甚至降低网络的探测效果。

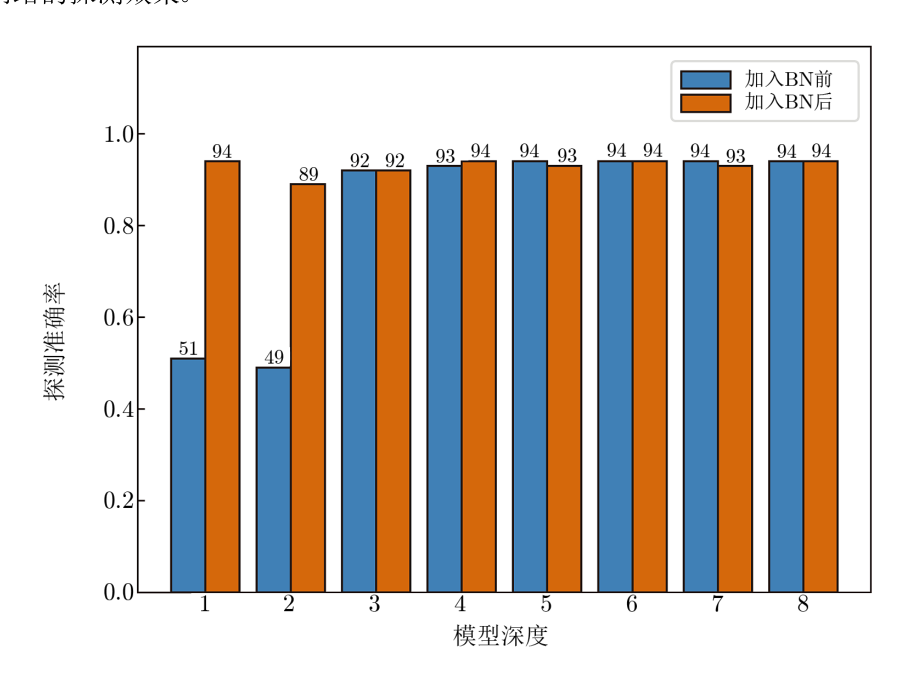


图 4 加入 BN 前后引力波探测准确率随模型深度的变化 (背景噪声为 H1 数据)

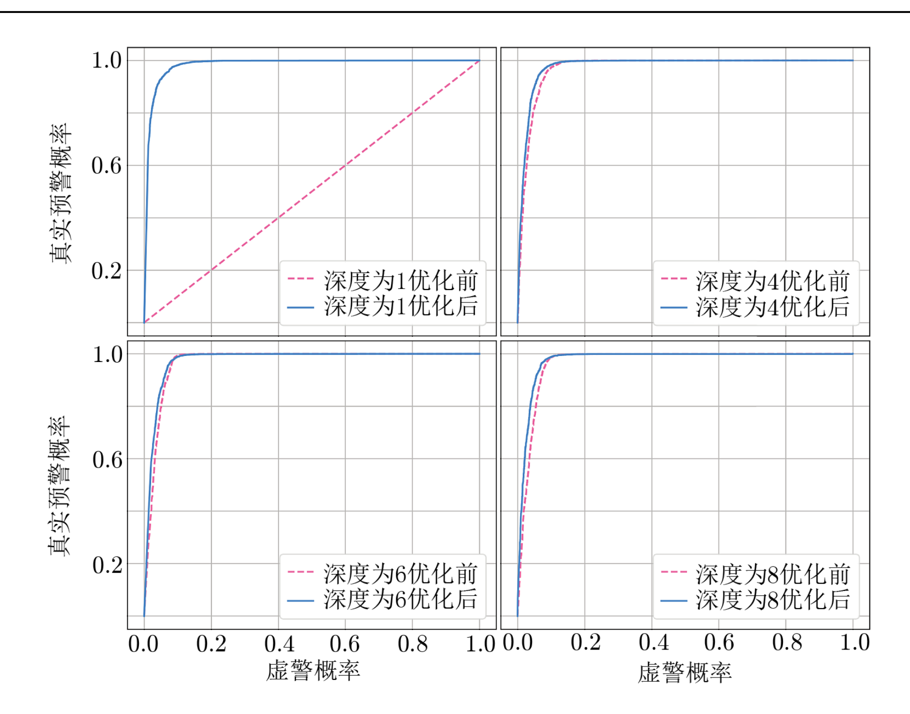


图 5　加入 BN 前后不同深度的模型在测试集上的 ROC 曲线 (背景噪声为 H1 数据)

为了探索 BN 层对不同深度模型的影响是否与不同探测器的工作状态有关，我们使用 L1 探测器噪声数据，并采用与上文相同方法对模型进行训练、测试。通过实验，我们得到与 H1 数据类似的结果。图 6 展示了在 L1 背景噪声下加入 BN 前后引力波探测准确率随模型深度的变化，由图可知，单卷积层神经网络在加入 BN 后探测准确率由 50% 提高到 91%。

5.3　加入 BN 层后网络模型泛化性研究

此外，为了验证加入 BN 层后网络模型的泛化能力，我们使用 H1 背景噪声数据训练的模型来对 L1 背景噪声合成的数据进行测试，具体测试结果如图 7 所示。我们发现 H1 数据训练的模型能够应用于 L1 数据的探测，但是泛化能力有所降低，如对于单层神经网络在 H1 数据中的探测准确率为 94%，在 L1 数据中的探测准确率降低为 89%。

6　结　论

引力波探测对研究黑洞的起源和演化等有重要的意义。本文对不同深度和宽度超参数空间的卷积神经网络应用于引力波探测进行了研究，为了排除训练集规模的影响，我们采用四种规模的训练集进行训练，分别是 4 800, 20 000, 35 000, 50 000。通过实验我们发现，模型每层卷积核数量越多 (宽度越宽)，最优深度就越小。可以通过综合调整模型的深度和宽度来优化探测模型的性能。不经过 BN 提升技术优化的单层和双层模型的引力波探测效果很差，在加入 BN 优化后，单层和双层卷积神经网络引力波探测效果大大提升，从 50% 左右

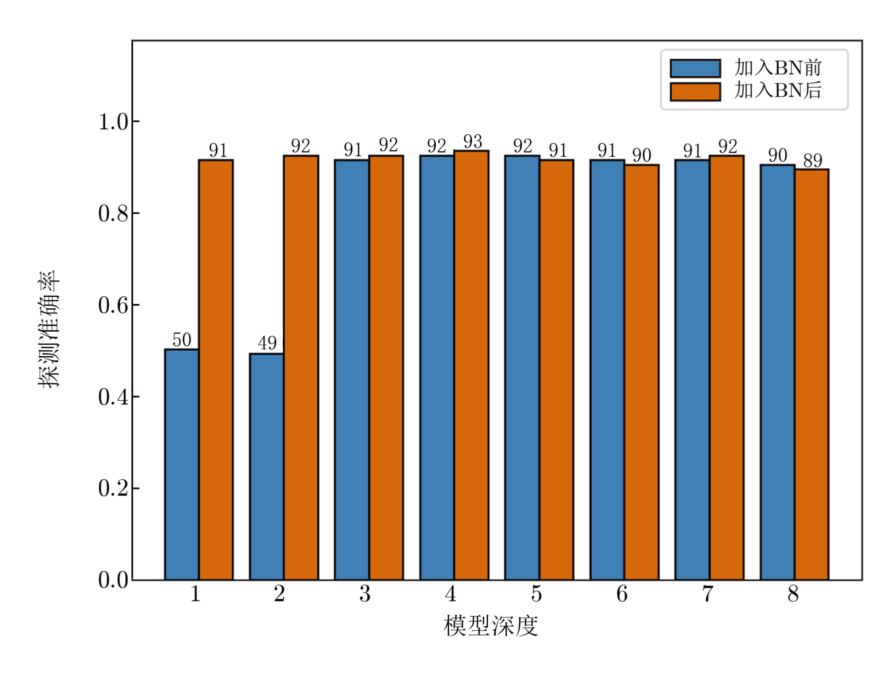


图 6 加入 BN 前后引力波探测准确率随模型深度的变化 (背景噪声为 L1 数据)

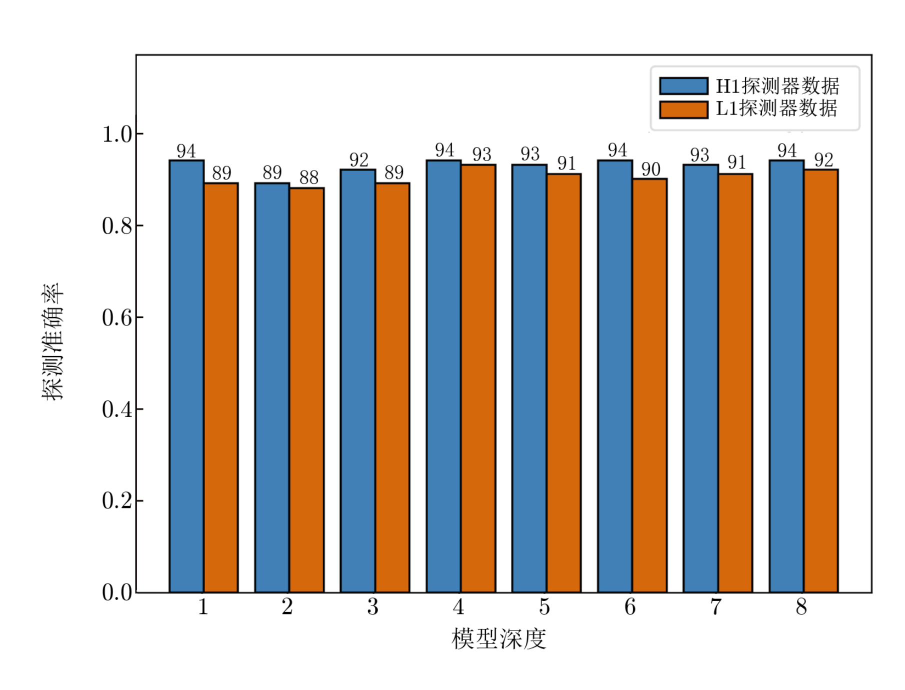


图 7 H1 数据训练模型对 L1 数据泛化性能测试结果

提升为 90% 以上。本文的结果为匹配滤波模板数量的压缩提供了潜在的新方法，匹配滤波后通过 BN 层和全连接层也许能够大大减少匹配模板数量。由于多层卷积神经网络模型的全连接层前加入 BN 优化的探测准确率可能降低，因此 BN 层所在的深度能够影响探测效果。

参考文献：

[1] Abbott B P, Abbott R, Abbott T D, et al. Phys Rev Lett, 2016, 116(6): 1102
[2] Abbott B P, Abbott R, Abbott T D, et al. Phys Rev Lett, 2017, 119(16): 1101
[3] Goldstein A, Veres P, Burns E, et al. ApJL, 2017, 848(2): L14
[4] Shoemaker I M, Murase K. Phys Rev D, 2018, 97(8): 083013
[5] Savchenko V, Ferrigno C, Kuulkers E, et al. ApJL, 2017, 848(2): L15
[6] Gao H. Sci. China Phys Mech, 2018, 61(5): 1
[7] Auclair P, Blanco-Pillado J J, Figueroa D G, et al. JCAP, 2020, 2020(04): 034
[8] Luo J, Chen L S, Duan H Z, et al. Classical Quant Grav, 2016, 33(3): 5010
[9] Hu W R, Wu Y L. Natl Sci Rev, 2017, 4(5): 685
[10] 王小鸽, ERIC Lebigot, 都志辉, 等. 天文学进展, 2016, 34(01): 50
[11] Nitz A H, Dal Canton T, Davis D, et al. Phys. Rev. D, 2018, 98(2): 4050
[12] Sachdev S, Caudill S, Fong H, et al. arXiv:1901.08580, 2019
[13] Hooper S, Chung S K, Luan J, et al. Phys Rev D, 2012, 86(2): 4012
[14] Vajente G, Huang Y, Isi M, et al. Phys Rev D, 2020, 101(4): 2003
[15] Gabbard H, Williams M, Hayes F, et al. Phys Rev Lett, 2018, 120(14): 1103
[16] George D, Huerta E A. Phys Rev D, 2018, 97(4): 4039
[17] George D, Huerta E A. Phys Lett B, 2018, 778: 64
[18] Gebhard T D, Kilbertus N, Harry I, et al. Phys Rev D, 2019, 100(6): 3015
[19] Wang H, Wu S, Cao Z, et al. Phys Rev D, 2020, 101(10): 4003
[20] Li X R, Yu W L, Fan X L, et al. Front Phys-Beijing, 2020, 15(5): 1
[21] Luo H M, Lin W b, Chen Z C, et al. Front Phys-Beijing, 2020, 15(1): 4601
[22] Deighan D S, Field S E, Capano C D, et al. arXiv:2010.04340, 2020
[23] Xia H, Shao L, Zhao J, et al. Phys Rev D, 2021, 103(2): 4040
[24] Wei W, Khan A, Huerta E A, et al. Phys Lett B, 2021, 812: 6029
[25] Shen H, Huerta E A, Zhao Z. arXiv:1903.01998, 2019
[26] Fan X L, Li J, Li X, et al. Sci China Phys Mech, 2019, 62(6): 1
[27] Chatterjee C, Wen L, Vinsen K, et al. Phys Rev D, 2019, 100(10): 3025
[28] Chua A, Vallisneri M. Phys Rev Lett, 2020, 124(4): 1102
[29] Shen H, George D, Huerta E A, et al. ICASSP, 2019: 3237
[30] Wei W, Huerta E A. Phys Lett B, 2020, 800: 135081
[31] Dreissigacker C, Sharma R, Messenger C, et al. Phys Rev D, 2019, 100(4): 4009
[32] Iess A, Cuoco E, Morawski F, et al. MLST, 2020, 1(2): 5014
[33] Edwards M C. Phys Rev D, 2021, 103(2): 4025
[34] Chen M N, Zhong Y H, Feng Y, et al. Sci China Phys Mech, 2020, 63(12): 1
[35] Goodfellow I, Bengio Y, Courville A. Deep learning[M]. Cambridge: MIT press, 2016: 331

Research on the Influence of Network Structure Complexity on Deep Learning for Gravitational Wave Detection

MA Cun-liang[1], ZHAN Chao[1], JIA Ming-zhen[1], HE Guan-sheng[2],
LI Wei-jun[2], YI Jian-bing[1]

(1. Jiangxi University of Science and Technology, Ganzhou 341000, China; 2. University of South China, Hengyang 421200, China)

Abstract: Deep learning for gravitational wave detection has been a hot spot in recent years. The matched filtering method can be regarded as a neural network with a single convolutional layer, and the template is stored in the convolution kernel. By increasing the depth of the model, similar detection effects can be obtained while the parameter number is greatly reduced. We study the deep learning gravitational wave detection models with different convolution kernel sizes, convolution kernel number, and convolution layer number. In addition, we investigate the detection effect of the model with batch normalization before the fully connected layer, and find that the detection accuracy of the model with a single convolutional layer increase from about 50% to more than 90% when the batch normalization is applied. A potential new method is provided for the compression of the number of matched filtering templates. By adding BN and full connection layer after matching filter, the number of matching templates may be greatly reduced. The generalization ability of the optimized CNN models on different background noise is also investigated. We find that the model trained by the data with H1 background noise can be used to detect the data with L1 background noise, but the accuracy is slightly reduced.

Key words: binary black hole merger; gravitational wave detection; deep learning